arXiv:1303.7255v1 [math.FA] 28 Mar 2013

OPTIMAL TRANSPORTATION OF PROCESSES WITH INFINITE KANTOROVICH DISTANCE. INDEPENDENCE AND SYMMETRY.

ALEXANDER V. KOLESNIKOV AND DANILA A. ZAEV

Abstract. We consider probability measures on $\mathbb{R}^\infty$ and study natural analogs of optimal transportation mappings for the case of infinite Kantorovich distance. Our examples include 1) quasi-product measures, 2) measures with certain symmetric properties, in particular, exchangeable and stationary measures. It turns out that the existence problem for optimal transportation is closely related to various ergodic properties. We prove the existence of optimal transportation for a certain class of stationary Gibbs measures. In addition, we establish a variant of the Kantorovich duality for the Monge–Kantorovich problem restricted to the case of measures invariant with respect of actions of compact groups.

1. Introduction

We prove the existence of optimal transportation mappings for certain classes of measures on $\mathbb{R}^\infty$. The optimal transportation mappings in finite-dimensional spaces can be constructed as solutions to the (quadratic) Monge–Kantorovich problem. Given a couple of probability measures μ and ν on $\mathbb{R}^d$ with Lebesgue densities, the corresponding optimal transportation mapping $T\colon \mathbb{R}^d \to \mathbb{R}^d$ gives a minimum to the functional

$$T \to \int \|x - T(x)\|^2 \, d\mu$$

among all mappings $T\colon \mathbb{R}^d \to \mathbb{R}^d$ transforming μ onto ν (here $\|\cdot\|$ is the standard Euclidean norm). It turns out that T has the form $T(x) = \nabla\varphi(x)$, where φ is a convex function.

The standard existence proof relies on the existence of the solution to the following problem for measures: find minimum of the functional

$$W_2^2(\mu,\nu) = \inf\Big\{ \int \|x-y\|^2 \, dm\colon m \in P(\mu,\nu)\Big\},$$

on the space $P(\mu,\nu)$ of probability measures with fixed projections: $Pr_x m = \mu, Pr_y m = \nu$. This problem is called the Monge–Kantorovich problem. Having a solution m to the Monge–Kantorovich problem, one can easily reconstruct the

Key words and phrases. Monge–Kantorovich problem, optimal transportation, Kantorovich duality, Gaussian measures, Gibbs measures, exchangeability, stationarity, ergodicity, transportation inequalities, entropy, and Kullback-Leibler distance.

This study was carried out within "The National Research University Higher School of Economics" Academic Fund Program in 2012-2013, research grant No. 11-01-0175. The first author was supported by RFBR projects 10-01-00518, 11-01-90421-Ukr-f-a, and the program SFB 701 at the University of Bielefeld. The second author was partially supported by AG Laboratory NRU-HSE, RF government grant, ag. 11.G34.31.0023.

desired mapping T. It turns out that the optimal measure m is supported on the graph of T:

$$m(\Gamma) = 1, \quad \text{where} \quad \Gamma = \{(x, T(x)), \ x \in \mathbb{R}^d\}.$$

The functional $W_2(\mu, \nu)$ is a distance in the space of probability measures. In what follows we call it the Kantorovich distance.

Another well-known fact which will be used throughout the paper is the following relation called the Kantorovich duality:

$$W_2(\mu, \nu) = J(\varphi, \psi),$$

where

$$J(\varphi, \psi) = \inf_{\varphi,\psi} \Bigl\{ \int \Bigl(\varphi(x) - \frac{x^2}{2}\Bigr) \ d\mu + \int \Bigl(\psi(y) - \frac{y^2}{2}\Bigr) \ d\nu, \quad \varphi(x) + \psi(y) \geq \langle x, y \rangle \Bigr\},$$

where the supremum is taken over couples of integrable Borel functions $\varphi(x), \psi(y)$. Note that the function φ in the dual problem coincides with the potential generating the transportation mapping: $T = \nabla\varphi$.

The mapping T exists under quite broad assumptions. It is sufficient that μ and ν have densities and admit finite second moments (see [16]). In this case the existence of φ and T follows immediately from the existence of the solution to the dual Kantorovich problem. More on optimal transportation can be found in [1], [6], [16].

The situation in the infinite-dimensional case is still not well-understood. The main reason for this is the fact that natural norms associated with measures are infinite almost everywhere. The archetypical example is given by the Cameron–Martin norm of a Gaussian measure. Nevertheless, for certain couples of measures the transportation problem has a natural formulation and a unique solution.

Example 1.1. Let $\gamma = \prod_{i=1}^{\infty} \gamma_i = \prod_{i=1}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}} \ dx_i$ be the standard Gaussian product measure on $\mathbb{R}^\infty$ and $H = l^2$, $\|x\|_H^2 = \sum_{i=1}^{\infty} x_i^2$ be the corresponding Cameron–Martin space. More generally, one can consider any abstract Wiener space.

The optimal transportation problem is well-understood for the case of measures μ and ν which are absolutely continuous with respect to γ. The most general results were obtained in [11] (another approach has been developed in [12]). In particular, for any given probability measure $f \cdot \gamma$ there exists a transportation mapping $T(x) = x + \nabla\varphi(x)$ minimizing the cost

$$\int \|T(x) - x\|_{l^2}^2 \ d\gamma$$

and transforming γ onto $f \cdot \gamma$, provided $\int f \log f \ d\gamma < \infty$. Analogously, there exists a transportation mapping transforming $f \cdot \gamma$ onto γ.

It is known (this follows from the so-called Talagrand transportation inequality) that under assumption $\int f \log f \ d\gamma < \infty$ the Kantorovich distance between γ and $f \cdot \gamma$ is finite

$$W_2^2(\gamma, f \cdot \gamma) = \int \|T(x) - x\|_{l^2}^2 \ d\gamma < \infty.$$

In particular, $\nabla\varphi(x) \in l^2$ for γ-almost all x. More on optimal transportation on the Wiener space, the corresponding Monge–Ampére equation, regularity issues, and

transportation on other infinite-dimensional spaces can be found in [4], [5], [7], [8], [10], and [9].

We state now the central problem of this paper.

Problem 1.2. Let μ and ν be two probability measures on $\mathbb{R}^\infty$. When does there exist a transportation mapping T transforming μ onto ν which is "optimal" for the cost function $c(x,y) = \|x-y\|^2_{l^2}$?

Note that **we don't assume finiteness of the Kantorovich distance** between the measures. Of course, it makes impossible in general to understand T as a solution to a certain minimization problem. Nevertheless, we have many good candidates to be called "optimal transportation" in many particular cases. The following example motivates our study.

Example 1.3. 1) Let $\mu = \prod_{i=1}^\infty \mu_i(dx_i)$, $\nu = \prod_{i=1}^\infty \nu_i(dx_i)$ be product probability measures. Assume that μ_i, ν_i have densities. Then there exists a mass transportation mapping T taking μ onto ν which has the form

$$T(x) = (T_1(x_1), \cdots, T_i(x_i), \cdots),$$

where $T_n(x_n)$ is the one-dimensional optimal transportation transforming μ_i onto ν_i.

2) Let us consider the Gaussian measure μ obtained from the standard Gaussian measure γ by a linear mapping $T(x) = Ax$ with A symmetric and positive. It is known (and can be obtained from the law of large numbers) that γ and μ are mutually singular even in the simplest case $A = 2 \cdot \mathrm{Id}$. T is "optimal" because it is linear and given by a positive symmetric operator. Heuristically,

$$T(x) = \frac{1}{2}\nabla\langle Ax, x\rangle.$$

It is clear that in both cases T cannot be obtained as a minimizer of a functional of the type $\int \|T(x) - x\|^2_{l^2}\ d\mu$.

We fix the standard basis $\{e_i\}$, $e_i = (\delta_{ij})$ in $\mathbb{R}^\infty$. Denote by $\mathbb{R}^n$ the subspace of $\mathbb{R}^\infty$ generated by $\{e_1, \cdots, e_n\}$ and by P_n the orthogonal projection onto $\mathbb{R}^n$.

In this paper we consider two (eventually, non-equivalent) definitions of the Monge–Kantorovich optimal mappings in the infinite-dimensional case:

- D1) limits of finite-dimensional optimal mappings,
- D2) (in the symmetric case) solutions to the classical Monge problem for another (finite) cost function constrained to a set of symmetric measures.

Almost everywhere in this paper we use approach D1). Let us briefly explain D2).

It is possible to give a meaning to the Monge–Kantorovich optimization problem if we restrict ourselves to a certain class of symmetric measures. In this paper we consider two types of symmetry: exchangeable measures (invariant with respect to finite permutations of coordinates) and stationary measures on $\mathbb{R}^{\mathbb{Z}}$ (invariant with respect to shifts of coordinates). Note that $\|x-y\|^2_{l^2}$ is symmetric with respect to both types of symmetry. More generally, let G be a group of linear operators which acts on $X = Y = \mathbb{R}^\infty$ and $X \times Y$: $x \to gx$, $(x,y) \to (gx, gy)$, $g \in G$ and preserves the cost function $c(x,y)$. We assume that every basic vector e_j can be obtained from every other e_i by the action of this group: there exists $g \in G$ such that $e_i = ge_j$. Note that under these assumptions all the coordinates are identically distributed. This leads us to the following definition: given G-invariant marginals μ

and ν we call π an optimal solution to the Monge–Kantorovich problem if π solves the Monge–Kantorovich problem

$$\int (x_1 - y_1)^2 \ d\pi \to \min$$

among all measures which are invariant with respect to G. If there exists a mapping T such that its graph $\Gamma = \{x, T(x)\}$ satisfies $m(G) = 1$, we say that T is an optimal transportation mapping transforming μ onto ν.

Remark 1.4.
1) In fact, we will use definition D1) throughout the paper (more precisely, Definition 1.6). See, however, Section 6.
2) The existence of a solution to the symmetric Monge–Kantorovich problem can be established by standard compactness arguments.
3) The corresponding optimal transportation (if exists) must preserve G-invariant sets. This observation allows us to construct the following counter-example to the existence Problem 1.2. See also Example 6.4.

Example 1.5. Let $\mu = \gamma$ be the standard Gaussian measure on $\mathbb{R}^\infty$ and

$$\nu = \frac{1}{2}(\gamma + \gamma_2)$$

be the average of γ and its homothetic image $\gamma_2 = \gamma \circ S^{-1}$, where $S(x) = 2x$. There is no any mass transportation T of μ to ν which preserves "rotational invariance" or even exchangeability of a set (i.e., if A is invariant with respect to cylindrical rotations, then $T(A)$ is invariant too). Indeed, any mapping of such a type must have the form $T(x) = g(x)(x_1, x_2, \cdots) = g(x) \cdot x$, where g is invariant with respect to any "rotation", in particular, with respect to any coordinate permutation. But any function g of this type is constant γ-a.e. This is a corollary of the Hewitt–Savage $0-1$ law. It is clear that there is no any mass transportation of this type for the given target measure.

There is a general principle behind this simple example. Recall that a measure μ is called ergodic with respect to a group action G, if for every G-invariant set A one has either $\mu(A) = 1$ or $\mu(A) = 0$. It follows directly from the definition that *there is no any bijective mass transportation T transforming μ onto ν, such that $T(A)$ is G-invariant for any G-invariant set A, provided μ is ergodic but ν is not.*

Definition 1.6. We say that a measurable mapping T transforming μ onto ν is optimal if the measure $\pi = \mu \circ (x, T(x))^{-1}$ on the graph

$$\{(x, T(x))| \quad x \in \mathrm{supp}(\mu)\} \subset \mathbb{R}^\infty \times \mathbb{R}^\infty$$

of T can be obtained as a weak limit of probability measures π_n such that 1) the support of π_n is contained in $\mathbb{R}^n \times \mathbb{R}^n$, 2) π_n is a solution of a finite-dimensional Kantorovich problem with quadratic cost $\sum_{i=1}^n |x_i - y_i|^2$.

Let us summarize the main results obtained in this paper. All of them are applicable to a partial case when $\nu = \gamma$, where γ is the standard Gaussian measure (or, more generally, ν is a uniformly log-concave product measure). This very special case of the target measure is important for applications (see, for example, [15]). In Section 3 we give some general sufficient conditions for the existence of transportation mappings. We prove existence in the following cases:

C1) μ is a quasi-product measure, i.e. μ has a density with respect to some product measure (Section 4),

C2) μ is exchangeable (Section 6),

C3) μ is stationary (Section 7).

The transportation in the quasi-product case C1) can be viewed as a perturbation of the "diagonal" transportation described in Example 1.3. More precisely, if μ and ν have densities with respect to (different) product measures P and Q, then the optimal transportation T transforming μ onto ν has the form

$$T = T_0 + \tilde{T},$$

where T_0 is the transportation of P onto Q described in Example 1.3 and $\tilde{T}$ is "small" compared to T_0. This result generalizes the results on the Wiener space obtained is [11], [12].

We use in C2) a De Finetti-type decomposition theorem, representing exchangeable measures as averages of countable powers of one-dimensional measures:

$$\mu = \int m^{\infty}(B)\Pi_{\mu}(dm), \quad \nu = \int m^{\infty}(B)\Pi_{\nu}(dm),$$

where m belongs to the space $\mathcal{P}(\mathbb{R})$ of Borel probablity measures on $\mathbb{R}$ and Π_{μ}, Π_{ν} are probability measures on $\mathcal{P}(\mathbb{R})$. We show that the problem of existence of an optimal transportation (in the sense of D2)) transforming μ onto ν is reduced (in a sense) to the optimal transportation problem for Π_{μ}, Π_{ν} with the squared Kantorovich distance on $\mathcal{P}(\mathbb{R})$ as the cost function.

The proof in C3) follows the ideas from [12]. We apply a Talagrand-type estimate of the L^2-distance between transportation mappings via the relative entropy of the corresponding measures. For any probability measures $\mu = e^{-V}\ dx$, $\nu = e^{-W}\ dx$ on $\mathbb{R}^d$ and the corresponding optimal transportation mappings T_{μ}, T_{ν}, taking μ, ν onto the standard Gaussian measure on $\mathbb{R}^d$, respectively, the following estimate holds:

$$\mathrm{Ent}_{\nu}\Bigl(\frac{\mu}{\nu}\Bigr) = \int \log \frac{d\mu}{d\nu}\ d\mu \geq \frac{1}{2} \int \|\nabla T_{\mu} - T_{\nu}\|^2\ d\mu.$$

The main example of Section 6 is given by a Gibbs measure on the lattice $\mathbb{R}^{\mathbb{Z}}$ with the following formal shift-invariant Hamiltonian:

$$H = \sum_{i=-\infty}^{\infty} V(x_i) + W(x_i, x_{i+1}).$$

The existence results for such measures can be found in [2].

In addition, we establish a variant of the Kantorovich duality for the Monge–Kantorovich problem restricted to the space of measures invariant with respect of actions of a compact group (Section 5). Unfortunately, we don't know so far, what is an adequate generalization of this statement for non-compact groups (which is the most interesting case).

2. Some preliminary estimates

Let μ and ν be probability measures on $\mathbb{R}^d$ and $T(x) = \nabla\varphi(x)$ be the optimal transportation transforming μ onto ν. Let us denote by μ_v the images of μ under the shifts $x \mapsto x + v$, $v \in \mathbb{R}^d$.

It will be assumed throughout that μ_v have densities with respect to μ:

$$\frac{d\mu_v}{d\mu} = e^{\beta_v}.$$

Lemma 2.1. *For every $p, q \ge 1$ with $\frac{1}{p} + \frac{1}{q} = 1$, $\varepsilon \ge 0$, and $e \in \mathbb{R}^d$*

$$\int |\varphi(x+te) - \varphi(x)|^{1+\varepsilon} \quad d\mu \le t^{1+\varepsilon} \| \, |\langle x, e\rangle|^{1+\varepsilon} \|_{L^p(\nu)} \cdot \sup_{0 \le s \le t} \|e^{\beta_{se}}\|_{L^q(\mu)}.$$

$$\int \big(\varphi(x+te) - \varphi(x) - t\partial_e \varphi(x)\big) \ d\mu \le t \|\langle x, e\rangle\|_{L^p(\nu)} \cdot \sup_{0 \le s \le t} \|e^{\beta_{se}} - 1\|_{L^q(\mu)}.$$

Proof. One has $\varphi(x+te) - \varphi(x) = \int_0^t \partial_e \varphi(x+se) \ ds$. Hence

$$\begin{aligned}
&\int |\varphi(x+te) - \varphi(x)|^{1+\varepsilon} \ d\mu \le t^{\varepsilon} \int \int_0^t |\partial_e \varphi|^{1+\varepsilon}(x+se) \ ds \ d\mu \\
&= t^{\varepsilon} \int_0^t \Big[\int |\partial_e \varphi|^{1+\varepsilon} e^{\beta_{se}} \ d\mu \Big] \ ds \le t^{1+\varepsilon} \| |\partial_e \varphi|^{1+\varepsilon} \|_{L^p(\mu)} \cdot \sup_{0 \le s \le t} \|e^{\beta_{se}}\|_{L^q(\mu)} \\
&= t^{1+\varepsilon} \| \, |\langle x, e\rangle|^{1+\varepsilon} \|_{L^p(\nu)} \cdot \sup_{0 \le s \le t} \|e^{\beta_{se}}\|_{L^q(\mu)}.
\end{aligned}$$

Applying the same arguments one gets

$$\begin{aligned}
&\int \big(\varphi(x+te) - \varphi(x) - t\partial_e \varphi(x)\big) \ d\mu = \int \int_0^t \big(\partial_e \varphi(x+se) - \partial_e \varphi(x)\big) \ ds \ d\mu \\
&= \int \Big[\int_0^t (e^{\beta_{se}} - 1) \ ds \Big] \partial_e \varphi(x) \ d\mu \le t^{\frac{1}{p}} \|\partial_e \varphi\|_{L^p(\mu)} \Big[\int \int_0^t |e^{\beta_{se}} - 1|^q \ ds \ d\mu \Big]^{\frac{1}{q}}.
\end{aligned}$$

The desired estimate follows from the the change of variables formula and trivial uniform bounds. $\square$

We recall that a probability measure μ on $\mathbb{R}^d$ is called log-concave if it has the form $e^{-V} \cdot \mathcal{H}^k|_L$, where $\mathcal{H}^k$ is the k-dimensional Haussdorff measure, $k \in \{0, 1, \cdots, d\}$, L is an affine subspace, and V is a convex function. We call a measure μ uniformly log-concave (more precisely, K-uniformly log-concave measure) if $\frac{1}{Z} e^{K|x|^2} \cdot \mu$ is a log-concave measure for some $K > 0$ and a suitable renormalization factor Z. It is well-known (C. Borell) that the projections and conditional measures of log-concave measures are log-concave. The same holds for uniformly log-concave measures. We can extend this notion to the infinite-dimensional case. Namely, we call a probability measure μ on a locally convex space X log-concave (K-uniformly log-concave with $K > 0$) if its images $\mu \circ l^{-1}$, $l \in X^*$ under linear continuous functionals μ are all log-concave (K-uniformly log-concave with $K > 0$).

Throughout the paper we apply the following estimate (see [12], [13]).

Proposition 2.2. *Let m be a K-uniformly log-concave probability measure with some $K > 0$. Then for any couple of probability measures $\mu = e^{-V} \ dx$, $\nu = e^{-W} \ dx$ and the corresponding optimal mappings $\nabla\varphi_{\mu}$, $\nabla\varphi_{\nu}$, transforming μ, ν onto m respectively, one has the following estimate*

$$\mathrm{Ent}_{\nu}\Big(\frac{\mu}{\nu}\Big) = \int \log \frac{d\mu}{d\nu} \ d\mu = \int (W - V) \ d\mu \ge \frac{K}{2} \int \big(\nabla\varphi_{\mu} - \nabla\varphi_{\nu}\big)^2 \ d\mu.$$

The quantity $\mathrm{Ent}_{\nu}\Big(\frac{\mu}{\nu}\Big)$ is called the relative entropy or Kullback-Leibler distance between μ and ν.

In addition, we will apply the following elementary Lemma.

Lemma 2.3. *Assume that a sequence $\{T_n\}$ of measurable mappings $T_n\colon \mathbb{R}^\infty \to \mathbb{R}^\infty$ converges to a mapping T in the following sense: for every e_i $\lim_n \langle T_n, e_i\rangle = \langle T, e_i\rangle$ in measure with respect to μ. Then the measures $\{\mu \circ T_n^{-1}\}$ converge weakly to $\mu \circ T^{-1}$.*

3. Sufficient conditions for existence

We consider a couple of Borel probability measures μ and ν on $\mathbb{R}^\infty$, where $\mathbb{R}^\infty$ is the space of all real sequences: $\mathbb{R}^\infty = \prod_{i=1}^\infty \mathbb{R}_i$. We deal with the standard coordinate system $x = (x_1, x_2, \cdots, x_n, \cdots)$ and the standard basis vectors $e_i = (\delta_{ij})$. The projection on the first n coordinates will be denoted by P_n: $P_n(x) = (x_1, \cdots, x_n)$. We use notations $\|x\|$, $\langle x, y\rangle$ for the Hilbert space norm and inner product: $\|x\| = \sum_{i=1}^\infty x_i^2$, $\langle x, y\rangle = \sum_{i=1}^\infty x_i y_i$. We use notation $\mathbb{E}_\mu^n$ for the conditional expectation with respect to μ and the σ-algebra generated by $x_1, \cdots, x_n$. For any product measure $P = \prod_{i=1}^\infty p_i(x_i)\ dx_i$ its projection $P_n = P \circ P_n^{-1}$ has the form $\prod_{i=1}^n p_i(x_i)\ dx_i$ and the projection $(f \cdot P) \circ P_n^{-1} = f_n \cdot P_n$ of the measure $f \cdot P$ satisfies $f_n = \mathbb{E}_P^n f$. Everywhere below we agree that every cylindrical function $f = f(x_1, \cdots, x_n)$ can be extended to $\mathbb{R}^\infty$ by the formula $x \to f_n(P_n x)$.

It will be assumed throughout the paper that the shifts of μ along any vector $v = te_i$ are absolutely continuous with respect to μ:

$$\frac{d\mu_v}{d\mu} = e^{\beta_v}.$$

In Section 3, moreover, the following assumption holds.

Assumption (A). For every basic vector $e = e_i$ there exist $p \geq 1$, $q \geq 1$, satisfying $\frac{1}{p} + \frac{1}{q} = 1$, and $\varepsilon > 0$ such that

$$\int |\langle x, e\rangle|^{(1+\varepsilon)p}\ d\nu < \infty$$

and

$$p(t) = \sup_{0 \leq s \leq t} \int |e^{\beta_{se}} - 1|^q\ d\mu$$

satisfies $\lim_{t\to 0} p(t) = 0$.

Let $\mu_n = \mu \circ P_n^{-1}(x)$, $\nu_n = \nu \circ P_n^{-1}(y)$ be the projections of μ, ν. For every $v = te_i$ let us set

$$\frac{d(\mu_n)_v}{d\mu_n} = e^{\beta_v^{(n)}}.$$

It is easy to check that the projections of μ, ν satisfy Assumption (A).

Lemma 3.1. *For every $n \in \mathbb{N}$ and every $e = e_i$ one has*

$$\int |\langle P_n(x), e\rangle|^p\ d\nu_n \leq \int |\langle x, e\rangle|^p\ d\nu, \quad \int |e^{\beta_e^{(n)}} - 1|^q\ d\mu_n \leq \int |e^{\beta_e} - 1|^q\ d\mu.$$

Proof. The first estimate is trivial. To prove the second one, let us note that $e^{\beta_v^{(n)}} = \mathbb{E}_\mu^n e^{\beta_v}$. The claim follows from the Jensen inequality and convexity of the function $t \to |t-1|^q$. $\square$

We denote by π_n the optimal transportation plan for the couple (μ_n, ν_n). Let $\varphi_n(x)$ and $\psi_n(y)$ solve the dual Kantorovich problem. Let us recall that $\nabla\varphi_n$ ($\nabla\psi_n$) is the optimal transportation mapping sending μ to ν (ν to μ). One has

$$\varphi_n(x) + \psi_n(y) \geq \langle P_n x, P_n y\rangle$$

for every x, y. The equality is attained on the support of π_n. In particular,

$$\varphi_n(x) + \psi_n(\nabla\varphi_n(x)) = \langle P_n x, \nabla\varphi_n(x)\rangle.$$

It is easy to check that $\{\pi_n\}$ is a tight sequence. By the Prokhorov theorem one can extract a weakly convergent subsequence $\pi_{n_k} \to \pi$. Note that π_n **is not** the projection of π.

In what follows we will pass several time to subsequences and use for the new subsequences the same index n again, with the agreement that n takes values in another infinite set $\mathbb{N}' \subset \mathbb{N}$. Let us fix unit vectors e_i, e_j for some $i, j \in \mathbb{N}$ and consider the following sequence of non-negative functions:

$$F_n(x, y, t, s) = \varphi_n(x + te_i) + \psi_n(y + se_j) - \langle P_n(x + te_i), P_n(y + se_j)\rangle$$

with $n > i, n > j$.

Lemma 3.2. *There exists a $L^{1+\varepsilon}(\pi)$-weakly convergent subsequence*

$$\varphi_{n_k}(x + te_i) - \varphi_{n_k}(x) \to U(x).$$

The following relation holds for the limiting function $U(x)$:

$$\Big|\int U(x)\ d\mu - t\int \langle y, e_i\rangle\ d\nu\Big| \le Ctp(t).$$

Proof. Taking into account that $\int F_n(x, y, 0, 0)\ d\pi_n = 0$, one obtains

$$\int F_n(x, y, t, 0)\ d\pi_n = \int F_n(x, y, t, 0)\ d\pi_n - \int F_n(x, y, 0, 0)\ d\pi_n \ge 0.$$

Note that the right-hand side equals

$$\int (F_n(x, y, t, 0) - F_n(x, y, t, 0))\ d\pi_n = \int \big[\varphi_n(x + te_i) - \varphi_n(x) - t\langle y, e_i\rangle\big]\ d\pi_n.$$

Taking into account that the projection of π_n onto X coincides with μ_n and φ_n depends on the first n coordinates, one finally obtains that for $n > i$ the latter is equal to

$$\int \big[\varphi_n(x+te_i) - \varphi_n(x)\big]\ d\mu - t\int \langle y, e_i\rangle\ d\nu = \int \big[\varphi_n(x+te_i) - \varphi_n(x) - t\partial_{e_i}\varphi_n(x)\big]\ d\mu$$

It follows from Lemma 2.1, Lemma 3.1 and Assumption (A) that

$$\Big|\int F_n(x, y, t, 0)\ d\pi_n\Big| \le Ctp(t). \tag{1}$$

Since φ_n depends on a finite number of coordinates, one has $\int |\varphi_n(x + te_i) - \varphi_n(x)|^{1+\varepsilon}\ d\mu = \int |\varphi_n(x + te_i) - \varphi_n(x)|^{1+\varepsilon}\ d\mu_n$. Hence by Lemma 2.1

$$U_n(x) = \varphi_n(x + te_i) - \varphi_n(x) \in L^{1+\varepsilon}(\mu)$$

and, moreover, $\sup_n \|U_n\|_{L^{1+\varepsilon}(\mu)} < \infty$. Thus there exists function $U \in L^{1+\varepsilon}(\mu)$ such that for some subsequence n_k

$$\varphi_{n_k}(x + te_i) - \varphi_{n_k}(x) \to U(x)$$

weakly in $L^{1+\varepsilon}(\mu)$. Passing to the limit we obtain from (1) that

$$\Big|\int U(x)\ d\mu - t\int \langle y, e_i\rangle\ d\nu\Big| \le Ctp(t).$$

$\square$

Lemma 3.3. *Assume that $F_n(x,y,0,0) \to 0$ in measure with respect to π. Then*

$$U(x) - t\langle y, e_i\rangle \geq 0$$

for π-almost all (x,y).

Proof. Note that

$$\bigl[\varphi_n(x+te_i)-\varphi_n(x)-t\langle y,e_i\rangle\bigr]+F_n(x,y,0,0) = \varphi_n(x+te_i)+\psi_n(y)-\langle P_n y, P_n(x+te_i)\rangle$$

is a non-negative function for every n. Since $F_n(x,y,0,0) \to 0$ in measure, there exists a subsequence (denoted again by F_n) which converges to zero π-almost everywhere. Since $f_n = \varphi_n(x+te_i) - \varphi_n(x) - t\langle y, e_i\rangle$ converges to $f = U(x) - t\langle y, e_i\rangle$ weakly in $L^{1+\varepsilon}(\pi)$, one can assume (passing again to a subsequence) that $\frac{1}{N}\sum_{n=1}^{N} f_n \to f$ π-a.e. Since $f_n + F_n \geq 0$, this implies that $f \geq 0$ π-a.e. $\square$

Proposition 3.4. *Assume that there exists a sequence of continuous functions*

$$f_n(x_1,\cdots,x_n), g_n(y_1,\cdots,y_n) \in L^1(\pi_n)$$

such that $G_n = f_n(x) + g_n(y) - \sum_{i=1}^n x_i y_i$ has the following properties:

1) $G_n \geq 0$,
2) $G_n \leq G_m, \quad \forall\ n \leq m, x, y \in \mathbb{R}^m$,
3) $\sup_n \int G_n \, d\pi_n < \infty$.

Then $F_n(x,y,0,0) \to 0$ in $L^1(\pi)$.

Proof. We start with the identity $\int F_n(x,y,0,0)\ d\pi_n = 0$ and rewrite it in the following way:

$$0 = \int (\varphi_n - f_n)\ d\mu + \int (\psi_n - g_n)\ d\nu + \int \bigl(f_n(x) + g_n(y) - \sum_{i=1}^n x_i y_i\bigr)\ d\pi_n. \tag{2}$$

Since φ_n, ψ_n are defined up to a constant, one can assume that $\int (\psi_n - g_n)\ d\nu = 0$. Thus $-\int(\varphi_n - f_n)\ d\mu = \int\bigl(f_n(x) + g_n(y) - \sum_{i=1}^n x_i y_i\bigr)\ d\pi_n$. It follows from 2) and 3) that both sides have limits. It follows from the weak convergence $\pi_n \to \pi$ and the monotonicity property 2) that for every k

$$\begin{aligned}\underline{\lim}_n \int \bigl(f_n(x) + g_n(y) - \sum_{i=1}^n x_i y_i\bigr)\ d\pi_n &\geq \underline{\lim}_n \int \bigl(f_k(x) + g_k(y) - \sum_{i=1}^k x_i y_i\bigr)\ d\pi_n \\ &= \int \bigl(f_k(x) + g_k(y) - \sum_{i=1}^k x_i y_i\bigr)\ d\pi.\end{aligned}$$

Hence

$$\underline{\lim}_n \int \bigl(f_n(x) + g_n(y) - \sum_{i=1}^n x_i y_i\bigr)\ d\pi_n \geq \lim_k \int \bigl(f_k(x) + g_k(y) - \sum_{i=1}^k x_i y_i\bigr)\ d\pi,$$

where the limit in the right-hand side exists, because the sequence is monotone. Hence we get from (2)

$$0 \geq \lim_n \int (\varphi_n - f_n)\ d\mu + \lim_n \int \bigl(f_n(x) + g_n(y) - \sum_{i=1}^n x_i y_i\bigr) d\pi.$$

Taking into account that $\int g_n \, d\pi = \int g_n \, d\nu = \int \psi_n \, d\nu = \int \psi_n \, d\pi$, we obtain

$$
\begin{aligned}
0 &\geq \lim_n \int (\varphi_n - f_n)(x) \, d\mu + \lim_n \int \big(f_n(x) + g_n(y) - \sum_{i=1}^n x_i y_i\big) d\pi \\
&= \lim_n \Big(\int (\varphi_n(x) + \psi_n(y) - \sum_{i=1}^n x_i y_i) \, d\pi \Big) \geq 0.
\end{aligned}
$$

The proof is complete. $\square$

Finally, we obtain a sufficient condition for the existence of an optimal mapping in the infinite-dimensional case.

Proposition 3.5. *Assume that $F_n(x, y, 0, 0) \to 0$ in measure with respect to π. Then there exists a mapping $T \colon \mathbb{R}^\infty \to \mathbb{R}^\infty$ such that*

$$T(x) = y$$

for π-almost all (x, y).

Proof. Let us fix e_i and choose a sequence of numbers $t_n \to 0$. We get from Lemma 3.2 and Lemma 3.3 that there exist π-a.e. nonnegative functions $U_{t_n}(x) - t_n \langle y, e_i \rangle$ with $\int \big(U_{t_n}(x) - t_n \langle y, e_i \rangle\big) \, d\pi = o(t_n)$. Hence, $\lim_{t_n \to 0} \int \big(\frac{U_{t_n}(x)}{t_n} - \langle y, e_i \rangle\big) \, d\pi = 0$. Taking into account that $\frac{U_{t_n}(x)}{t_n} - \langle y, e_i \rangle \geq 0$ for π-almost all (x, y), we conclude that $\frac{U_{t_n}(x)}{t_n}$ converges μ-a.e. and in $L^1(\mu)$ to a function $u_i(x)$ satisfying $u_i(x) - \langle y, e_i \rangle \geq 0$ and $\int (u_i(x) - \langle y, e_i \rangle) \, d\pi = 0$. Clearly, $u(x) = \langle y, e_i \rangle$ for π-almost all (x, y). Repeating these arguments for every $i \in \mathbb{N}$, we get the claim. $\square$

4. Quasi-product case

In what follows we consider two product measures

$$P = \prod_{i=1}^{\infty} p_i(x_i) \, dx_i$$

and

$$Q = \prod_{i=1}^{\infty} q_i(x_i) dx_i.$$

Measures which have densities with respect to a product measure will be called quasi-product measures.

Let

$$T(x) = (T_1(x_1), \cdots, T_n(x_n), \cdots)$$

be the infinite-dimensional the diagonal transportation mapping, where $T_i(x_i)$ transforming $p_i(x_i) dx_i$ onto $q_i(x_i) \, dx_i$. Clearly, T takes P onto Q. The inverse mapping $S = T^{-1}$ has the same structure:

$$S(x) = (S_1(x_1), \cdots, S_n(x_n), \cdots).$$

Theorem 4.1. *Let $\mu = f \cdot P$ and $\nu = g \cdot Q$ be probability measures satisfying the Assumption (A) of the previous section. Assume, in addition, that*

1) *there exists $K > 0$ such that every ν_i is K-uniformly log-concave;*

2) *there exists $M > 0$ such that*

$$S_i'(x_i) \le M;$$

uniformly in i and x;

3) *there exist c, C such that $0 < c \le g \le C$;*

4)

$$f \log f \in L^1(P).$$

Then there exists a optimal transportation mapping T transforming μ onto ν.

Remark 4.2. It follows from Caffarelli's contraction theorem (see [14]) that assumption 2) is satisfied if $(-\log p_i(x_i))'' \ge C_0$, $(-\log q_i(x_i))'' \le C_1$ for some $C_0, C_1 > 0$ and every i. Of course, there exist many other examples when this assumption is satisfied.

Proof. Consider the finite-dimensional projections $\mu_n = f_n \cdot P_n$, $\nu_n = g_n \cdot Q_n$, where $P_n = \prod_{i=1}^n p_i(x_i)\ dx_i$, $Q_n = \prod_{i=1}^n q_i(x_i)\ dx_i$. Here f_n and g_n are the conditional expectations of f, g with respect to P, Q and the σ-algebra $\mathcal{F}_n$. Recall that $\nabla \varphi_n$ is the optimal transportation of μ_n to ν_n. Let $u_i(x_i), v_i(y_i) = u_i^*$ be the one-dimensional potentials associated to the mappings T_i, S_i, respectively: $T_i = u_i', S_i = v_i'$. Note that $\tilde{T}_n = (T_1, \cdots, T_n)$ maps P_n onto Q_n and $\nabla \varphi_n$ maps $\frac{f_n}{g_n(\nabla \varphi_n)} \cdot P_n$ onto Q_n.

According to Proposition 2.2 one has the following estimate:

$$\frac{K}{2} \int |\tilde{T}_n - \nabla \varphi_n|^2 \frac{f_n}{g_n(\nabla \varphi_n)} dP_n \le \int \log\Bigl(\frac{f_n}{g_n(\nabla \varphi_n)}\Bigr) \frac{f_n}{g_n(\nabla \varphi_n)} dP_n. \tag{3}$$

Applying uniform bounds on g one easily gets that $c \le g_n \le C$ uniformly in n. It follows from the Jensen inequality that

$$\int f_n \log f_n dP \le \int f \log f dP. \tag{4}$$

Clearly, (3), (4) imply that

$$\sup_n \int |\tilde{T}_n - \nabla \varphi_n|^2 f_n dP_n = \sup_n \int |\tilde{T}_n - \nabla \varphi_n|^2 d\mu < \infty.$$

We complete the proof by applying Proposition 3.5 to the sequences of functions $\sum_{i=1}^n u_i(x_i)$, $\sum_{i=1}^n v_i(y_i)$. We need to estimate $\sum_{i=1}^n \int (u_i(x_i) + v_i(y_i) - x_i y_i)\ d\pi_n$. Taking into account that π_n is supported on the graph of $\nabla \varphi_n$, and the relation $u_i(x_i) + v_i(T_i(x)) = x_i T_i(x)$ we obtain that the latter equals to

$$\begin{aligned}
&\int \bigl(u_i(x_i) + v_i(\partial_{x_i} \varphi_n) - x_i \partial_{x_i} \varphi_n(x)\bigr)\ d\mu_n \\
&= \int \Bigl[v_i(\partial_{x_i} \varphi_n(x)) - v_i(T_i(x)) - x_i(\partial_{x_i} \varphi_n(x) - T_i(x))\Bigr]\ d\mu_n \\
&= \int \Bigl[v_i(\partial_{x_i} \varphi_n(x)) - v_i(T_i(x)) - v_i'(T_i(x))(\partial_{x_i} \varphi_n(x) - T_i(x))\Bigr]\ d\mu_n \\
&\le C \int (\partial_{x_i} \varphi_n(x) - T_i)^2\ d\mu_n.
\end{aligned}$$

Here we use the uniform bound $v_i'' = S_i' \le M$. Finally, we obtain that

$$\sum_{i=1}^n \int (u_i(x_i) + v_i(y_i) - x_i y_i)\ d\pi_n \le M \int |\nabla \varphi_n - \tilde{T}_n|^2\ d\mu_n.$$

We have already shown that the right-hand side is uniformly bounded in n. The result follows from Proposition 3.5. □

5. Kantorovich duality in the class of measures with compact invariance group

In this section we start to study measures which are invariant under actions of some group.

We begin with the most favorable situation: compact spaces and groups. The result of this section will not be used in this paper, but it is of independent interest.

Let X, Y be compact metric spaces, G be a compact group with bijective continuous actions L^X and L^Y on X, Y respectively. The action L on the product space $X \times Y$ is defined as follows:

$$L_g(X \times Y) = L_g^X(X) \times L_g^Y(Y).$$

Let μ and ν be Borel probability measures which are invariant under the actions L^X and L^Y respectively: $\mu, \nu \in Inv_L \iff \mu \circ (L_g^X)^{-1} = \mu, \nu \circ (L_g^Y)^{-1} = \nu$. We fix a non-negative and lower-semicontinuous cost function $c\colon X \times Y \to \mathbb{R}$. Denote by Π the set of all non-negative Borel probability measures on $X \times Y$ with the property $\pi \in \Pi \iff \pi = \mu \circ (Pr_X)^{-1}$ and $\pi \circ (Pr_Y)^{-1} = \nu$.

The space of L_G-invariant continuous functions V_L is a vector subspace of the space C_b, so there exists factor space $W_L = C_b/V_L$ of functions with the property: $\int_{g \in G} (\hat{u} \circ L_g^{-1}) d\chi(g) = 0$, where $d\chi(g)$ is normalized Haar measure on G. It is not hard to verify, that $C_b = V_L \oplus W_L$, and we can uniquely decompose every function u from $C_b(X \times Y)$ into the sum of a L_G-invariant function $\bar{u}$ from V_L and a function $\hat{u}$ from W_L:

$$u = \bar{u} + \hat{u}$$

It is clear, that $\bar{u} = \int_{g \in G} (u \circ L_g^{-1}) d\chi(g) = Pr_{V_L}(u)$, and it is a continuous projection of u onto V_L.

In the theorem below we generalize the well-known Kantorovich duality for the case of G-invariant constraints.

Theorem 5.1. *In the setting described above the following identity holds:*

$$\inf_{\pi \in \Pi \cap Inv_L} \int_{X \times Y} c(x, y) d\pi = \sup_{\bar{\phi} + \bar{\psi} \le \bar{c}} \left(\int_X \bar{\phi}(x) d\mu + \int_Y \bar{\psi}(y) d\nu \right).$$

Proof. The proof is based on the Fenchel–Rockafellar duality theorem.

Theorem 5.2. (Fenchel-Rockafellar duality) *Let X be a Polish space and let Θ and Ω be convex functionals from $C_b(X)$ to $\mathbb{R} \cup \{+\infty\}$. Assume that Θ is continuous at some point. Let Θ^* and Ω^* be the Legendre–Fenchel transforms of Θ and Ω considered on the space of Radon measures $M := (C_b)^*$ on X and defined as follows:*

$$\Theta^*(\pi) = \sup_{u \in C_b} \left(\int_X u d\pi - \Theta(u) \right), \quad \Omega^*(\pi) = \sup_{u \in C_b} \left(\int_X u d\pi - \Omega(u) \right).$$

Then the following Kantorovich-type duality holds:

$$\inf_{u \in C_b} \big(\Theta(u) + \Omega(u) \big) = \sup_{\pi \in M} \big(-\Theta^*(-\pi^*) - \Omega^*(\pi^*) \big) \tag{5}$$

Let

$$\Theta(u) = \begin{cases} 0 & \text{if } u(x,y) \geq -c(x,y) \\ +\infty & \text{else} \end{cases}$$

$$\Omega(u) = \begin{cases} \int_X \phi(x)d\mu + \int_Y \psi(y)d\nu & \text{if } u(x,y) = \phi(x) + \psi(y) + \hat{\omega}(x,y) \\ +\infty & \text{else} \end{cases}$$

where $\hat{\omega}(x,y)$ is a function from $W_L(X \times Y)$. It can be checked that they are both convex and Θ are continuous at $u = 0$. Let us find their Legendre–Fenchel transforms:

$$\Theta^*(-\pi) = \sup_{u \in C_b(X\times Y)} \left(-\int_{X\times Y} u(x,y)d\pi; \;\; u(x,y) \geq -c(x,y) \right).$$

If the measure π is not nonnegative, then there exists $v \in C_b(X \times Y)$ such that $\int v d\pi > 0$. Then we can choose $u = \lambda v$, $\lambda \to \infty$ and see, that the supremum of our functional is $+\infty$. In the other, case when π is nonnegative, it's clearly that supremum is $\int c d\pi$. So:

$$\Theta^*(-\pi) = \begin{cases} \int_{X\times Y} c(x,y)d\pi & \text{if } \pi \in M^+(X \times Y) \\ +\infty & \text{else} \end{cases}$$

$$\begin{aligned} \Omega^*(\pi) = \sup_{u \in C_b(X\times Y)} \Big(\int_{X\times Y} u(x,y)d\pi - \int_X \phi(x)d\mu - \int_Y \psi(y)d\nu, \\ u(x,y) = \phi(x) + \psi(y) + \hat{\omega}(x,y) \Big) = \\ = \sup_{\psi,\phi,\hat{\omega}} \left(\int_{X\times Y} (\psi + \phi)d\pi + \int_{X\times Y} \hat{\omega}(x,y)d\pi - \int_X \phi(x)d\mu - \int_Y \psi(y)d\nu \right). \end{aligned}$$

If $\pi \notin \Pi$, then there exist $\phi_1, \psi_1 \in C_b(X \times Y)$, $\hat{\omega}_1 \in W_L(X \times Y)$ such that $\int_{X\times Y} (\psi_1 + \phi_1)d\pi - \int_X \phi_1(x)d\mu - \int_Y \psi_1(y)d\nu > 0$ and $\int_{X\times Y} \hat{\omega}_1 d\pi \geq 0$. Then the choice $\phi = \lambda\phi_1$, $\psi = \lambda\psi_1$ $\lambda \to \infty$ shows that the supremum of Ω^* is $+\infty$. Similarly, if $\pi \in \Pi$, but $\pi \notin Inv_L$, then there exists $\hat{\omega}_1 \in W_L(X\times Y)$ such that $\int_{X\times Y} \hat{\omega}_1 d\pi > 0$ and the other terms vanish. Again the choice $\hat{\omega} = \lambda\hat{\omega}_1$, $\lambda \to \infty$ shows, that the supremum is $+\infty$. Obviously, in the last case: $\pi \in \Pi \cap Inv_L$ the supremum of Ω^* is 0. Thus:

$$\Omega^*(\pi) = \begin{cases} 0 & \text{if } \pi \in \Pi \cap Inv_L \\ +\infty & \text{else.} \end{cases}$$

Calculate the right-hand side of (5):

$$\begin{aligned} \sup_{\pi \in M} \left(-\Theta^*(-\pi^*) - \Omega^*(\pi^*)\right) = \sup_{\pi \in \Pi \cap Inv_L} \int_{X\times Y} -c(x,y)d\pi = \\ = -\inf_{\pi \in \Pi \cap Inv_L} \int_{X\times Y} c(x,y)d\pi. \end{aligned}$$

In the left-hand side of the duality statement we have

$$\inf_{u \in C_b} \big(\Theta(u) + \Omega(u)\big) =$$

$$= \inf_{\phi,\psi,\hat{\omega}} \left(\int_X \phi(x) d\mu + \int_Y \psi(y) d\nu;\ \ \phi(x) + \psi(y) + \hat{\omega}(x,y) \geq -c(x,y) \right) =$$

$$= - \sup_{\phi,\psi,\hat{\omega}} \left(\int_X \phi(x) d\mu + \int_Y \psi(y) d\nu;\ \ \phi(x) + \psi(y) + \hat{\omega}(x,y) \leq c(x,y) \right).$$

We are going to use the fact that μ and ν are invariant measures under the actions L^X and L^Y of the group G. This implies that $\int_X \phi(x) d\mu + \int_Y \psi(y) d\nu = \int_X \bar{\phi}(x) d\mu + \int_Y \bar{\psi}(y) d\nu$. So, we get

$$- \sup_{\phi,\psi,\hat{\omega}} \left(\int_X \phi(x) d\mu + \int_Y \psi(y) d\nu;\ \ \bar{\phi}(x) + \bar{\psi}(y) + \hat{\phi} + \hat{\psi} + \hat{\omega} \leq \bar{c} + \hat{c} \right) =$$

$$= - \sup_{\bar{\phi},\bar{\psi},\hat{\phi},\hat{\psi},\hat{\omega}} \left(\int_X \bar{\phi}(x) d\mu + \int_Y \bar{\psi}(y) d\nu;\ \ \bar{\phi}(x) + \bar{\psi}(y) \leq \bar{c} + (\hat{c} - \hat{\omega} - \hat{\phi} - \hat{\psi}) \right)$$

Note that the maximizing functional does not depend on the W_L-parts of ϕ and ψ, thus we can choose $\hat{\phi}$ and $\hat{\psi}$ arbitrary. Hence $\tilde{c}\colon = \hat{c} - \hat{\omega} - \hat{\phi} - \hat{\psi}$ is just an arbitrary function from W_L. The inequality $\bar{\phi}(x) + \bar{\psi}(y) \leq \bar{c} + \tilde{c}$ holds pointwise, so we can act on it by any element $g \in G$:

$$(\bar{\phi} \circ L_g)(x) + (\bar{\psi} \circ L_g)(y) \leq (\bar{c} \circ L_g + \tilde{c} \circ L_g)(x,y) \iff$$
$$\bar{\phi}(x) + \bar{\psi}(y) \leq (\bar{c} + \tilde{c} \circ L_g)(x,y)$$

for any $x, y \in X \times Y$. So:

$$\bar{\phi}(x) + \bar{\psi}(y) \ \leq\ (\bar{c} + \tilde{c})(x,y) \quad \iff \quad \bar{\phi}(x) + \bar{\psi}(y) \ \leq\ \left(\bar{c} + \inf_{g \in G} (\tilde{c} \circ L_g) \right)(x,y)$$

It follows immediately from the definition of W_L that $\inf_{g \in G} (\tilde{c} \circ L_g) \leq 0$. Hence the supremum is reached at $\tilde{c} \equiv 0$ ($\hat{\omega} = \hat{c}$, $\phi = \bar{\phi}$, $\psi = \bar{\psi}$). Finally, we get:

$$- \sup_{\bar{\phi},\bar{\psi} \in V_L, \tilde{c} \in W_L} \left(\int_X \bar{\phi}(x) d\mu + \int_Y \bar{\psi}(y) d\nu;\ \ \bar{\phi}(x) + \bar{\psi}(y) \leq \bar{c} + \tilde{c} \right) =$$
$$- \sup_{\bar{\phi} + \bar{\psi} \leq \bar{c}} \left(\int_X \bar{\phi}(x) d\mu + \int_Y \bar{\psi}(y) d\nu \right)$$

Collecting everything together we get from (5) the required statement. □

6. Exchangeable measures

In this section we discuss the mass transportation of exchangeable measures. Recall that a probability measure is exchangeable if it is invariant with respect to any permutation of finite number of coordinates. Before we consider $\mathbb{R}^\infty$, let us make some remarks on the finite-dimensional case.

Let S_d be the group of all permutations of $\{1, \cdots, d\}$. This group acts on $\mathbb{R}^d$ as follows:

$$L_\sigma(x) = (x_{\sigma(1)}, x_{\sigma(2)}, \cdots, x_{\sigma(d)}), \quad \sigma \in S_d.$$

Let $\Gamma \subset S_d$ be any subgroup which acts transitively. The latter means that for every couple i, j there exists $\sigma \in \Gamma$ such that $\sigma(i) = j$.

Assume that the source and image measures are both invariant with respect to Γ. In this case the Kantorivich potential φ is also Γ-invariant: $\varphi = \varphi \circ L_\sigma$ for any $\sigma \in \Gamma$. Consequently, the optimal transportation $T = \nabla \varphi$ has the following property:

$$T = L^*_\sigma (T \circ L_\sigma) = L^{-1}_\sigma \circ T \circ L_\sigma.$$

Equivalently,

$$L_\sigma \circ T = T \circ L_\sigma.$$

The action of Γ can be extended to $\mathbb{R}^d \times \mathbb{R}^d$ as follows: $L_\sigma(x,y) = (L_\sigma x, L_\sigma y)$. The optimal transportation plan $\pi(dx,dy)$ is also Γ-invariant.

Now let $\sigma(i) = j$. One has

$$\begin{aligned}\int x_i T_i \ d\mu &= \int \langle e_i, x\rangle \langle e_i, T(x)\rangle \ d\mu = \int \langle L_\sigma e_i, L_\sigma x\rangle \langle e_i, L^*_\sigma (T(L_\sigma x)\rangle \ d\mu \\ &= \int \langle e_j, L_\sigma x\rangle \langle e_j, (T(L_\sigma x)\rangle \ d\mu = \int \langle e_j, x\rangle \langle e_j, T(x)\rangle \ d\mu = \int x_j T_j \ d\mu.\end{aligned}$$

Consequently,

$$W_2^2(\mu,\nu) = \int \|x - T(x)\|^2 \ d\mu = \sum_{i=1}^{d} \int (x_i - T_i(x))^2 \ d\mu = d \int (x_i - T_i(x))^2 \ d\mu, \ \ \forall i.$$

Conclusion: *The quadratic Monge–Kantorovich problem for Γ-invariant marginals is equivalent to the transportation problem for the cost $|x_1 - y_1|^2$ restricted to the set of Γ-invariant probability measures.*

We denote by S_∞ be the group of permutation $\mathbb{N}$ which permutes a finite number of coordinates. We consider its natural action on $\mathbb{R}^\infty$ defined by

$$\sigma(x) = (x_{\sigma(i)}), \quad x = (x_i) \in \mathbb{R}^\infty, \quad \sigma \in S_\infty.$$

In this section we consider measures μ and ν which are invariant with respect to any $\sigma \in S_\infty$:

$$\mu = \mu \circ \sigma^{-1}, \quad \nu = \nu \circ \sigma^{-1}.$$

The measures of this type are called exchangeable.

Example 6.1. Let m be a Borel probability measure on $\mathbb{R}$. Its countable power m^∞ is an exchangeable measure on $\mathbb{R}^\infty$.

The conclusion made above helps us to give a variational meaning to the transportation problem in the infinite-dimensional case.

Definition 6.2. Let μ and ν be exchangeable. Consider the set $\mathcal{P}_{S_\infty}$ of probability measures on $X \times Y, X = Y = \mathbb{R}^\infty$ which are invariant with respect to any mapping $(x,y) \to (L_\sigma(x), L_\sigma(y))$, $\sigma \in S_\infty$ and have fixed S_∞-invariant marginals μ and ν. We say that a measure $\pi \in \mathcal{P}_{S_\infty}$ is a solution to the quadratic Monge–Kantorovich problem if it gives the minimum to the functional

$$\mathcal{P}_{S_\infty} \ni m \to \int (x_1 - y_1)^2 \ dm. \tag{6}$$

Assume that there exists a measurable mapping $T \colon X \to Y$ such that $m(\{(x,T(x))\}) = 1$. Then we say that T is a optimal transportation of μ onto ν.

Clearly, a solution to the Monge–Kantorovich problem (6) exists provided $\int x_1^2\, d\mu < \infty, \int y_1^2\, d\nu < \infty$. The corresponding optimal mapping T (if exists) must commute with any L_σ. This means that for μ-almost all x

$$T \circ L_\sigma(x) = L_\sigma \circ T(x). \tag{7}$$

The set of exchangeable measures is described in the following generalization of the classical De Finetti theorem (see, e.g., [3], Theorem 10.10.19).

Theorem 6.3. *Let $\mathcal{P}$ be the space of Borel probability measures on $\mathbb{R}$ equipped with the weak topology. Then for every Borel exchangeable μ on $\mathbb{R}^\infty$ there exists a probability measure Π on $\mathcal{P}$ such that*

$$\mu(B) = \int m^\infty(B)\Pi(dm),$$

for every Borel $B \subset \mathbb{R}^\infty$.

The structure of mappings satisfying (7) is easy to describe. Assume first that μ is a product measure. Consider the function $T_1(x) = \langle T(x), e_1\rangle$ and fix the first coordinate x_1. Then the function $F\colon (x_2, x_3, \cdots) \to T_1(x)$ is invariant with respect to S_∞. Hence F is constant according by the Hewitt–Sawage $0-1$ law. Thus $T_1(x) = T_1(x_1)$ depends on x_1 only and the mapping T is diagonal: $(T_1(x_1), T_2(x_2), \cdots)$.

Example 6.4. Let μ_1, μ_2 be countable powers of two different one-dimensional measures. By the Kakutani dichotomy theorem they are mutually singular. There is no any mass transportation T of $\mu = \mu_1$ onto $\nu = \frac{1}{2}(\mu_1 + \mu_2)$ satisfying (7). Indeed, any T satisfying (7) must be diagonal, hence the measure $\mu \circ T^{-1}$ must be a product measure.

Thus, we see that the optimal transportation does not always exist. Let us find sufficient conditions for the existence. We consider a couple of exchangeable measures μ, ν and their mixture representations:

$$\mu = \int m^\infty\, d\Pi_\mu, \quad \nu = \int m^\infty\, d\Pi_\nu.$$

By the strong law of large numbers, for any Borel function f one has for m^∞-almost any x

$$\int f\, dm = \lim_n \frac{1}{n}\big(f(x_1) + \cdots + f(x_n)\big).$$

Let us choose a sequence of bounded continuous functions $\{f_i\}$ on $\mathbb{R}$ which is dense in $C([a,b])$ for any a, b and set

$$\mathcal{S}_m = \bigcap_{i=1}^\infty \Big\{x\colon \lim_n \frac{1}{n}\big(f_i(x_1) + \cdots + f_i(x_n)\big) = \int f_i\, dm\Big\}.$$

Clearly, $m^\infty(\mathcal{S}_m) = 1$, but $p^\infty(\mathcal{S}_m) = 0$ for $p \ne m$.

For any couple of measures m_1^∞, m_2^∞ we set

$$d^2(m_1^\infty, m_2^\infty) := W_2^2(m_1, m_2).$$

Recall that $W_2^2(m_1, m_2)$ is the squared Kantorovich distance between m_1, m_2.

Let T be a transportation mapping which 1) transforms μ onto ν, 2) commutes with S_∞, 3) gives minimum to the functional $\int |T_1(x) - x_1|^2\, d\mu$ among of the mappings satisfying 1) and 2). It follows from the considerations above that T

must be diagonal on any set $\mathcal{S}_m$ (up to a m^∞-measure zero and for Π_μ-almost all m). This means that for Π_μ-almost all m one has

$$T|_{\mathcal{S}_m} = T_{m,F(m)}, \quad m^\infty\text{−a.e.},$$

where $F\colon \mathcal{P}(\mathbb{R}) \to \mathcal{P}(\mathbb{R})$ is a Borel mapping and $T_{m,F(m)}$ is the diagonal optimal transportation of m^∞ onto $F^\infty(m)$. Computing the transportation cost

$$\int |T_1 - x_1|^2 \, d\mu = \int W_2^2(m, F(m)) d\Pi_\mu = \int d^2(m^\infty, F^\infty(m)) d\Pi_\mu$$

we get that F must be a optimal transportation of Π_μ onto Π_ν for the cost function $(m,p) = W_2^2(m,p)$.

Taking into account all the remarks above, we arrive at the following conclusion.

Proposition 6.5. *Let the mapping T satisfy the following assumptions*

1) $\nu = \mu \circ T^{-1}$,
2) $L_\sigma \circ T = T \circ L_\sigma \quad \mu - a.e.$,
3) $\int |T_1(x) - x_1|^2 \, d\mu$ *is minimal among all mappings satisfying 1) and 2).*

Then up to a set of μ-zero measure $T|_{\mathcal{S}_m} = T_{m,F(m)}$, where F is a optimal transportation of Π_μ onto Π_ν for the cost function $(m,p) \to W_2^2(m,p)$ and $T_{m,F(m)}$ is the diagonal optimal transportation of m^∞ onto $F^\infty(m)$.

Remark 6.6. The mapping F can be considered as a kind of "factorization" of T. Of course, as we have already seen, T and F do not always exist.

Proposition 6.5 does not give, however, any checkable sufficient conditions for the existence of a optimal transportation. The questions, whether the mapping T from Proposition 6.5 can be approximated by finite-dimensional optimal mappings seems to be non-trivial. In the rest of the section we give some constructive sufficient conditions for the existence of the optimal transportation of exchangeable measures, where the transportation is understood (as before) as a limit of finite-dimensional approximations (see Definition 1.6).

Recall that the projection $\mu \circ P_n^{-1}$ of μ onto the first n coordinates is denoted by μ_n. For a couple of numbers $m < n$ we denote by $P_{m,n}$ the projection onto the subspace generated by $\{e_{m+1}, \cdots, e_n\}$ and by $\mu_{m,n} = \mu \circ P_{m,n}^{-1}$ the corresponding image of μ

It will be assumed in the rest of this section that for any couple of numbers $n > m$ the projection μ_n is absolutely continuous with respect to $\mu_n \times \mu_{m,n}$, hence there exists a representation

$$\mu_n = \rho_{m,n} \cdot \mu_m \times \mu_{m,n} \tag{8}$$

and the same holds for ν:

$$\nu_n = d_{m,n} \cdot \nu_m \times \nu_{m,n}. \tag{9}$$

Theorem 6.7. *Assume that*

1) *the measures μ and ν are exchangeable;*
2) *all the projections μ_n, ν_n admit Lebesgue densities and the representation (8), (9);*
3) *there exists a sequence $C_m > 0, \varepsilon > 0$ such that for every $m, n \in \mathbb{N}$ one has*

$$\int \log \rho_{m,n} d\mu < C_m, \quad \int d_{m,n}^{-(1+\varepsilon)} d\nu < C_m$$

and

$$\lim_{m\to\infty} \frac{C_m}{m} = 0;$$

4) *the measure ν is K-uniformly log-concave for some $K > 0$.*

Then there exists a optimal transportation mapping T transforming μ onto ν.

Remark 6.8. The assumption 3) means, in particular, that the Kullback–Leibler distance between μ_n and $\mu_n \times \mu_{m\times n}$ is $o(m)$ uniformly in n.

Proof. Let T_m be the optimal transportation mapping transforming μ_m onto ν_m and $T_{m,n}$ be the optimal transportation mapping transforming $\mu_{m,n}$ onto $\nu_{m,n}$. Clearly,

$$\tilde{T}_{m,n}(x) = T_m(P_m x) + T_{m,n}(P_{m,n} x)$$

maps $\mu_m \times \mu_{m,n}$ onto $\nu_m \times \nu_{m,n}$. Using the representation $\nu_n = d_{m,n} \cdot \nu_m \times \nu_{m,n}$, we get that $\tilde{T}_{m,n}$ transforms

$$d_{m,n}(\tilde{T}_{m,n}) \cdot \mu_m \times \mu_{m,n}$$

onto ν_n.

By Proposition 2.2 one has the following estimate:

$$\mathrm{Ent}_{\rho_{m,n}(\tilde{T}_{m,n})\mu_m \times \mu_{m,n}} \Bigl(\frac{\mu_n}{\rho_{m,n}(\tilde{T}_{m,n})\mu_m \times \mu_{m,n}}\Bigr) \ge \frac{K}{2} \int \bigl\|\tilde{T}_{m,n} - T_n\bigr\|^2 \, d\mu_n.$$

This implies that

$$\int \log\Bigl(\frac{\rho_{m,n}}{d_{m,n}(\tilde{T}_{m,n})}\Bigr) \, d\mu_n \ge \frac{K}{2} \int \bigl\|T_m - P_m \circ T_n\bigr\|^2 \, d\mu_n = \frac{K}{2} \int \bigl\|T_m - P_m \circ T_n\bigr\|^2 \, d\mu.$$

Let us estimate the left-hand side of

$$\int \log\Bigl(\frac{\rho_{m,n}}{d_{m,n}(\tilde{T}_{m,n})}\Bigr) \, d\mu_n = \int \log \rho_{m,n} \, d\mu - \int \log d_{m,n}(\tilde{T}_{m,n}) \, d\mu_n.$$

The desired estimate for the first term follows immediately from the assumptions that $\int \log \rho_{m,n} \, d\mu \le C$. Let us apply the inequality

$$xy \le \frac{1}{\varepsilon}\Bigl(e^{\varepsilon x} + y \log y - y\Bigr),$$

which holds true for any x and any $y > 0$, $\varepsilon > 0$. One has

$$\begin{aligned}
-\int \log d_{m,n}(\tilde{T}_{m,n}) \, d\mu_n &= -\int \log d_{m,n}(\tilde{T}_{m,n}) \rho_{m,n} \, d\mu_m \times \mu_{m,n} \\
&\le \frac{1}{\varepsilon}\Bigl(\int \frac{1}{d^{\varepsilon}_{m,n}(\tilde{T}_{m,n})} \, d\mu_m \times \mu_{m,n} + \int (\rho_{m,n} \log \rho_{m,n} - \rho_{m,n}) \, d\mu_m \times \mu_{m,n}\Bigr) \\
&= \frac{1}{\varepsilon}\Bigl(\int \frac{1}{d^{\varepsilon}_{m,n}} \, d\nu_m \times \nu_{m,n} + \int (\log \rho_{m,n} - 1) \, d\mu_n\Bigr) \\
&= \frac{1}{\varepsilon}\Bigl(\int \frac{1}{d^{1+\varepsilon}_{m,n}} \, d\nu_n + \int (\log \rho_{m,n} - 1) \, d\mu\Bigr). \\
&= \frac{1}{\varepsilon}\Bigl(\int \frac{1}{d^{1+\varepsilon}_{m,n}} \, d\nu + \int (\log \rho_{m,n} - 1) \, d\mu\Bigr).
\end{aligned}$$

Applying assumption 3) we finally obtain that there exists a sequence $c_m = o(m)$ such that

$$\frac{K}{2} \int \bigl\|T_m - P_m \circ T_n\bigr\|^2 \, d\mu_n = \frac{K}{2} \int \bigl\|T_m - P_m \circ T_n\bigr\|^2 \, d\mu < c_m.$$

Since μ and ν are exchangeable, the same holds for all projections μ_n, ν_n. This implies, in particular, that

$$\int \left\|T_m - P_m \circ T_n\right\|^2 \, d\mu = m \int \langle T_m - T_n, e_i\rangle^2 \, d\mu$$

for every $1 \le i \le m$ and $n > m$. Hence $\int \langle T_m - T_n, e_i\rangle^2 \, d\mu \le \frac{2c_m}{mK}$. Passing to a $L^2(\mu)$-weakly convergent subsequence $\langle T_{n_k}, e_i\rangle \to T_i$ we obtain in the limit

$$\int (\langle T_m, e_i\rangle - T_i)^2 \, d\mu \le \frac{2c_m}{mK}.$$

Clearly, this gives $\lim_m \langle T_m, e_i\rangle = T_i$ in $L^2(\mu)$. It follows from Lemma 1.3 that $T = \sum_{i=1}^{\infty} T_i \cdot e_i$ is the desired mapping. $\square$

All assumptions of Theorem 6.7 are easy to check, excepting 3). Let us give a simple example, where it can be checked.

Example 6.9. Let μ be a finite convex combination of measures which are countable products of measures with Lebesgue densities. Then there exists $C > 0$ such that $\sup_n \int \log \rho_{m,n} d\mu < C$.

Proof. Let μ on $\mathbb{R}^\infty$ be a finite convex combination of product measures (finite mixture). This means that it has the form $\mu = \sum \lambda_i \mu_i$, $\sum \lambda_i = 1$, where every μ_i is a product measure. The same holds for every projection

$$\mu \circ P_n^{-1} = \sum \lambda_i (\mu_i \circ P_n^{-1}).$$

Set:

$$p_i = \frac{d\mu_i \circ P_m^{-1}}{dx}, \quad q_i = \frac{d\mu_i \circ P_{m,n}^{-1}}{dx}.$$

Then $\rho_{m,n} = \frac{d\mu_n}{d\mu_m \times \mu_{m,n}}$ can be expressed as

$$\rho_{m,n} = \frac{\sum \lambda_i p_i q_i}{(\sum \lambda_i p_i)(\sum \lambda_j q_j)}.$$

We claim that $\int \log \rho_{m,n} d\mu < C$ for some constant C independent of n, m. Indeed,

$$\rho_{m,n} = \frac{\sum \lambda_i p_i q_i}{(\sum \lambda_i p_i)(\sum \lambda_j q_j)} = \frac{\sum \lambda_i p_i q_i}{\sum \lambda_i^2 p_i q_i + \sum_{i \ne j} \lambda_i \lambda_j p_i q_j} \le \frac{\sum \lambda_i p_i q_i}{\sum \lambda_i^2 p_i q_i}.$$

Using the trivial estimate

$$\frac{\sum \lambda_i p_i q_i}{\sum \lambda_i^2 p_i q_i} \le \frac{1}{\inf_i (\lambda_i)}$$

we finally get

$$\int \log \rho_{m,n} d\mu = \int \log \frac{\sum \lambda_i p_i q_i}{(\sum \lambda_i p_i)(\sum \lambda_j q_j)} d\mu \le$$
$$\le \int \log \frac{1}{\inf_i (\lambda_i)} d\mu = -\log \inf_i (\lambda_i) < C.$$

Let us remark that this argument works well only with finite mixtures. $\square$

7. Stationary Gibbs measures

In this section we study stationary Gibbs measures. Unlike the previous section we identify our space $\mathbb{R}^\infty$ not with $\mathbb{R}^\mathbb{N}$, but with $\mathbb{R}^\mathbb{Z}$. This means that we allow negative coordinate indices: $x_0, x_{-1}, x_{-2}, \cdots$.

Set

$$E_n = \mathrm{span}\{e_i, -n \le i \le n\}$$

and

$$E_{m,n} = \mathrm{span}\{e_i, e_j, -n \le i < -m, m < j \le n\}.$$

The corresponding orthogonal projections will be denoted by $P_n, P_{m,n}$ accordingly.

Let $\sigma_n \colon E_n \to E_n$ be the cyclical shift

$$\sigma_n \colon (x_{-n}, x_{-(n-1)}, \cdots, x_{-1}, x_0, x_1, \cdots, x_n) \to (x_n, x_{-n}, x_{-(n-1)}, \cdots, x_{-1}, x_0, \cdots, x_{n-1}).$$

In the limit $n \to \infty$ we obtain the standard (Bernoulli) shift

$$\sigma \colon x = (x_i) \to (x_{i-1}).$$

Definition 7.1. A probability measure μ is called stationary if it is invariant with respect to σ: $\mu \circ \sigma^{-1} = \mu$.

Throughout the section the following assumptions hold.

1) The measure μ is a weak limit
$$\mu = \lim_n \mu_n,$$
where every μ_n is a σ_n-invariant measure on E_n with everywhere positive Lebesgue density and finite second moments;
2) For every $m < n$ there exists a probability measure $\mu_{m,n}$ on $E_{m,n}$ such that the relative entropy (the Kullback-Leibler distance) between $\mu_m \times \mu_{m,n}$ and μ_n is uniformly bounded in n:
$$\int \log\Bigl(\frac{d\mu_n}{d(\mu_m \times \mu_{m,n})}\Bigr) d\mu_n < C_m$$
with C_m satisfying
$$\lim_m \frac{C_m}{m} = 0;$$
3) For every power l of the cyclical shift σ_m the measure μ has density with respect to $\mu \circ (\sigma_m^l)^{-1}$. Moreover,
$$e^{u_{m,l}} = \frac{d\mu}{d\mu \circ (\sigma_m^l)^{-1}}$$
satisfies $\sup_{m,l} \|e^{u_{m,l}}\|_{L^{2+\delta}(\mu)} < \infty$ for some $\delta > 0$;
4) Every μ_n is absolutely continuous with respect to $\mu \circ P_n^{-1}$: $\mu_n = \rho_n \cdot \mu \circ P_n^{-1}$ and, in addition,
$$\sup_n \int \rho_n^{2+\delta} \, d\mu < \infty$$
for some $\delta > 0$.

Remark 7.2. We note that 1) + 4) imply convergence of $\{\mu_n\}$ in a stronger sense. Namely, $\lim_n \int \varphi \, d\mu_n = \int \varphi \, d\mu$ for every cylindrical $\varphi \in L^2(\mu)$. Indeed, take a

continuous bounded cylindrical function $\tilde{\varphi}$ such that $\|\varphi - \tilde{\varphi}\|_{L^2(\mu)} < \varepsilon$. One has $\lim_n \int \varphi \, d\mu_n = \lim_n \int (\varphi - \tilde{\varphi}) \, d\mu_n + \int \tilde{\varphi} \, d\mu$. The claim follows from the estimate

$$\Bigl(\int |\varphi - \tilde{\varphi}| \, d\mu_n\Bigr)^2 \le \int (\varphi - \tilde{\varphi})^2 \, d\mu \cdot \int \rho_n^2 \, d\mu \le (\sup_n \int \rho_n^2 \, d\mu)\varepsilon^2$$

Remark 7.3. The idea of the proof of Theorem 7.4 below is the same as in Theorem 6.7. Applying the Talagrand-type inequality from Proposition 2.2 and the symmetric properties of measures we show L^2-convergence of the finite-dimensional approximations. The proof of Theorem 7.4 is much longer because we need to overcome the following technical difficulty: the projections of a stationary measure are not invariant with respect to cyclical shifts. Thus the situation is different to the exchangeable case, where we have stability under projections.

Theorem 7.4. *Let μ be a probability measure satisfying Assumptions 1)-4). Then there exists a optimal transportation mapping transforming μ onto the standard Gaussian measure on $\mathbb{R}^\infty$.*

Proof. Let us consider the n-dimensional optimal transportation T_n transforming μ_n onto the standard n-dimensional Gaussian measure γ_n. It follows from the σ_n-invariance of μ_n and γ_n that the mapping T_n is cyclically invariant:

$$\langle T_n \circ \sigma_n, e_i \rangle = \langle T_n, e_{i-1} \rangle, \quad \mu_n - \text{a.e.}$$

(with the convention $e_{n+1} = e_{-n}, e_{-n-1} = e_n$).

Let us fix a couple of numbers m, n with $n > m$. Let $T_{m,n}$ be the optimal transportation mapping transforming $\mu_{m,n}$ onto the standard Gaussian measure on $E_{m,n}$. We stress that T_m and $T_{m,n}$ depend on different collections of coordinates.

We extend T_m onto $\mathbb{R}^n$ in the following way:

$$T_m(x) = T_m(P_m x) + T_{m,n}(P_{m,n} x).$$

Clearly, T_m maps $\mu_m \times \mu_{m,n}$ onto the standard Gaussian measure on E_n. Applying Proposition 2.2 to the couple of mappings T_m, T_n, we get

$$\frac{1}{2} \int \|T_n - T_m\|^2 d\mu_n \le \int \log\Bigl(\frac{d\mu_n}{d(\mu_m \times \mu_{m,n})}\Bigr) d\mu_n. \tag{10}$$

Thus

$$\sum_{i=-m}^{m} \int \langle T_n - T_m, e_i \rangle^2 \, d\mu_n \le C_m \tag{11}$$

for every m, n, $m < n$.

Let us note that for every i one can extract a weakly convergent subsequence from a sequence of (signed) measures $\{\langle T_n, e_i \rangle \cdot \mu_n\}$. Indeed, for any compact set K

$$\Bigl(\int_{K^c} |\langle T_n, e_i \rangle| d\mu_n\Bigr)^2 \le \int |\langle T_n, e_i \rangle|^2 d\mu_n \cdot \mu_n(K^c) = \int x_i^2 \, d\gamma \cdot \mu_n(K^c).$$

Using the tightness of $\{\mu_n\}$ we get that $\{|\langle T_n, e_i \rangle| \cdot \mu_n\}$ is a tight sequence. In addition, note that for every continuous f

$$\lim_n \Bigl(\int f |\langle T_n, e_i \rangle| d\mu_n\Bigr)^2 \le \int x_i^2 \, d\gamma \cdot \int f^2 d\mu.$$

This implies that any limiting point of $\{\langle T_n, e_i \rangle \cdot \mu_n\}$ is absolutely continuous with respect to μ. Applying the diagonal method and passing to a subsequence one can

assume that convergence takes place for all i simultaneously. Consequently, there exists a subsequence $\{n_k\}$ and a measurable mapping T with values in $\mathbb{R}^\infty$ such that

$$\langle T_{n_k}, e_i\rangle \cdot \mu_{n_k} \to \langle T, e_i\rangle \cdot \mu$$

weakly in the sense of measures for every i. It is easy to check that the standard property of L^2-weak convergence holds also in this case:

$$\int \langle T, e_i\rangle^2 d\mu \le \underline{\lim}_k \int \langle T_{n_k}, e_i\rangle^2 d\mu_n = \int x_i^2 \, d\gamma = 1 \tag{12}$$

Finally, we pass to the limit in (11) (here we apply (12) and Remark 7.2) and obtain

$$\sum_{i=-m}^{m} \int \langle T - T_m, e_i\rangle^2 \, d\mu \le C_m. \tag{13}$$

Note that T commutes with the shift σ: $\langle T\circ\sigma, e_i\rangle = \langle T, e_{i-1}\rangle$. Indeed, for every bounded cylindrical φ one has

$$\int \varphi\langle T_n, e_{i-1}\rangle d\mu_n = \int \varphi\langle T_n(\sigma_n), e_i\rangle d\mu_n = \int \varphi(\sigma_n^{-1})\langle T_n, e_i\rangle d\mu_n = \int \varphi(\sigma^{-1})\langle T_n, e_i\rangle d\mu_n.$$

Here we use that $\varphi(\sigma_n^{-1}) = \varphi(\sigma^{-1})$ for sufficiently large values of n and the cyclical invariance of T_n. Passing to the limit in the n_k-subsequence one gets

$$\int \varphi\langle T, e_{i-1}\rangle d\mu = \int \varphi(\sigma^{-1})\langle T, e_i\rangle d\mu = \int \varphi\langle T\circ\sigma, e_i\rangle d\mu.$$

Hence $T\circ\sigma = \sigma\circ T$.

Using invariances, we get that

$$\begin{aligned}\int \langle T - T_m, e_i\rangle^2 \, d\mu &= \int \langle T\circ\sigma^l - T_m\circ\sigma_m^l, e_{i+l}\rangle^2 d\mu \\ &= \int \langle T\circ\sigma^l\circ(\sigma_m^l)^{-1} - T_m, e_{i+l}\rangle^2 e^{-u_{m,l}} d\mu\end{aligned}$$

(for $i, i+l \in [-m, m]$). Applying assumption 3) we get by the Cauchy–Bunyakovski inequality

$$\begin{aligned} C\int \langle T - T_m, e_i\rangle^2 \, d\mu &\ge \int e^{u_{m,l}} \, d\mu \int \langle T\circ\sigma^l\circ(\sigma_m^l)^{-1} - T_m, e_{i+l}\rangle^2 e^{-u_{m,l}} d\mu \\ &\ge \Bigl(\int |\langle T\circ\sigma^l\circ(\sigma_m^l)^{-1} - T_m, e_{i+l}\rangle| d\mu\Bigr)^2. \end{aligned} \tag{14}$$

We note that $\lim_m \sigma^l\circ(\sigma_m^l)^{-1}(x) = x$ for every x. Let us show that

$$\langle T\circ\sigma^l\circ(\sigma_m^l)^{-1}, e_i\rangle \to \langle T, e_i\rangle$$

in $L^1(\mu)$. To this end take a continuous bounded function $\tilde{T}_i$ such that $\int |\tilde{T}_i - \langle T, e_i\rangle|^2 \, d\mu \le \varepsilon^2$. Then

$$\begin{aligned}\int |\langle T\circ\sigma^l\circ(\sigma_m^l)^{-1} - T, e_i\rangle| \, d\mu &\le \int |\tilde{T}_i\circ\sigma^l\circ(\sigma_m^l)^{-1} - \tilde{T}_i| \, d\mu \\ &+ \int |\langle T\circ\sigma^l\circ(\sigma_m^l)^{-1}, e_i\rangle - \tilde{T}_i\circ\sigma^l\circ(\sigma_m^l)^{-1}| \, d\mu + \int |\langle T, e_i\rangle - \tilde{T}_i| \, d\mu.\end{aligned}$$

We will show that the first integral in the right-hand side tends to zero as $m\to\infty$ and the others are small (of the order ε). We prove the first statement only, for the second one the arguments are similar. Note that $\lim_m \tilde{T}_i\circ\sigma^l\circ(\sigma_m^l)^{-1} = \tilde{T}_i$

poinwise. It is sufficient to show that $\sup_m \int |\tilde{T}_i \circ \sigma^l \circ (\sigma^l_m)^{-1}|^{1+\varepsilon} \, d\mu < \infty$ for some $\varepsilon > 0$.

One easily gets the desired estimate:

$$\int |\tilde{T}_i \circ \sigma^l \circ (\sigma^l_m)^{-1}|^{1+\varepsilon} d\mu = \int |\tilde{T}_i \circ \sigma^l|^{1+\varepsilon} e^{u_{m,l}} d\mu \le \|\tilde{T}_i\|_{L^2(\mu)} \|e^{u_{m,l}}\|_{L^{2/(1-\varepsilon)}(\mu)} \le C$$

where $2/(1-\varepsilon) = 2+\delta$ and C is independent of m,.

Using convergence $\langle T \circ \sigma^l \circ (\sigma^l_m)^{-1}, e_k \rangle \to \langle T, e_k \rangle$ in $L^1(\mu)$ we get from (14) and (13) that for some C independent of m, l

$$C \cdot \underline{\lim}_m \int \langle T - T_m, e_i \rangle^2 \, d\mu \ge \underline{\lim}_m \Bigl(\int |\langle T - T_m, e_{i+l} \rangle| d\mu \Bigr)^2.$$

In particular, fixing some i_0 and setting $l = i_0 - i$, we get

$$C \cdot C_m > C \cdot \underline{\lim}_m \sum_{i=-m}^{m} \int \langle T - T_m, e_i \rangle^2 \, d\mu \ge \underline{\lim}_m m \cdot \Bigl(\int |\langle T - T_m, e_{i_0} \rangle| d\mu \Bigr)^2.$$

Applying the relation $\lim_m \frac{C_m}{m} = 0$, we get that $\underline{\lim} \Bigl(\int |\langle T - T_m, e_{i_0} \rangle| d\mu \Bigr)^2 = 0$ for every i_0. Hence one can extract a subsequence m_k such that $\langle T_{m_k}, e_i \rangle \to \langle T, e_i \rangle$ in $L^1(\mu)$ for every i (in what follows we denote this subsequence again by $\{T_n\}$). Moreover, it follows from (11) that one has convergence in $L^{2-\varepsilon}(\mu)$ for every $\varepsilon > 0$.

It remains to show that T maps μ onto γ. Fix a smooth Lipschitz function f which depends on a finite number of coordinates $(x_{-k}, \cdots, x_k)$. By the change of variables formula one has for every $n \ge k$

$$\int f(T_n) \, d\mu_n = \int f \, d\gamma.$$

In the other hand, let us approximate $P_k T$ by a mapping $\tilde{T}_\varepsilon \colon \mathbb{R}^\infty \to P_k(\mathbb{R}^\infty)$ such that $\|P_k(T) - \tilde{T}_\varepsilon\|_{L^{(2+\delta)^*}(\mu)} \le \varepsilon$ and every $\langle \tilde{T}_\varepsilon, e_j \rangle$ is smooth and bounded for every $-k \le j \le k$. One has

$$\begin{aligned} &| \int f(T_n) \, d\mu_n - \int f(\tilde{T}_\varepsilon) \, d\mu_n | \le \|f\|_{Lip} \int \|P_k(T_n) - \tilde{T}_\varepsilon\| \, d\mu_n \\ &\le C \|f\|_{Lip} \Bigl(\|P_k(T_n - T)\|_{L^{(2+\delta)^*}(\mu)} + \|P_k(T) - \tilde{T}_\varepsilon\|_{L^{(2+\delta)^*}(\mu)} \Bigr), \end{aligned}$$

where $C = \sup_n \|\rho_n\|_{L^{(2+\delta)^*}(\mu)}$, $\rho_n = \frac{d\mu_n}{d\mu \circ P_n^{-1}}$. Using $L^{(2+\delta)^*}(\mu)$-convergence $\langle T_n, e_i \rangle \to \langle T, e_i \rangle$, smoothness of $f(\tilde{T}_\varepsilon)$, and the weak convergence $\mu_n \to \mu$, we pass to the limsup in the inequality. We obtain

$$\overline{\lim}_n | \int f(T_n) \, d\mu_n - \int f(\tilde{T}_\varepsilon) \, d\mu | \le \varepsilon C \|f\|_{Lip}.$$

Choosing an appropriate sequence $\tilde{T}_\varepsilon \to T$ we get $\int f(T) \, d\mu = \lim_n \int f(T_n) \, d\mu_n = \int f \, d\gamma$. The proof is complete. $\square$

Below we apply Theorem 7.4 to Gibbs measures. We study a transportation of a Gibbs measure μ which can be formally written in the form

$$\mu = e^{-H(x)} dx,$$

where the potential H admits the following heuristic representation:

$$H(x) = \sum_{i=-\infty}^{+\infty} V(x_i) + \sum_{i=-\infty}^{+\infty} W(x_i, x_{i+1}).$$

Here V and W are smooth functions and $W(x,y)$ is symmetric: $W(x,y) = W(y,x)$. The existence of such measures was proved in [2].

Let us specify the assumptions about V and W below. These are a particular case of assumptions A1-A3 from [2].

1)
$$W(x,y) = W(y,x);$$

2) There exist numbers $J > 0$, $L \geq 1$, $N \geq 2$, $\sigma > 0$, and $A, B, C > 0$ such that
$$|W(x,y)| \leq J(1+|x|+|y|)^{N-1}, \ |\partial_x W(x,y)| \leq J(1+|x|+|y|)^{N-1}$$

3)
$$|V(x)| \leq C(1+|x|)^L, \quad |V'(x)| \leq C(1+|x|)^{L-1};$$

4) (coercitivity assumption)
$$V'(x) \cdot x \geq A|x|^{N+\sigma} - B.$$

Let us define the following probability measure on E_n:

$$\mu_n = \frac{1}{Z_n} \exp\Big(- \sum_{i=-n}^{n} \big(V(x_i) + W(x_i, x_{i+1})\big)\Big),$$

with the convention $x_{n+1} := x_{-n}$. Here Z_n is the normalizing constant.

Proposition 7.5. *The sequence μ_n admits a weakly convergent subsequence $\mu_{n_k} \to \mu$ satisfying the assumptions of Theorem 7.4.*

Proof. It was proved in Theorem 3.1 of [2] that any sequence of probability measures

$$\tilde{\mu}_n = c_n e^{-H_n} dx_{-n} \cdots dx_n,$$

where H_n is obtained from H by fixing a boundary condition $\tilde{x}$

$$H_n = \sum_{i=-n}^{n} V(x_i) + \sum_{i=-n}^{n-1} W(x_i, x_{i+1}) + W(\tilde{x}_{-(n+1)}, x_{-n}) + W(x_n, \tilde{x}_{n+1}),$$

has a weakly convergent subsequence $\tilde{\mu}_{n_k} \to \tilde{\mu}$. In addition (see [2]), μ satisfies the following a priori estimate: for every $\lambda > 0$

$$\sup_{k \in \mathbb{Z}} \int \exp(\lambda |x_k|^N) \ d\tilde{\mu} < \infty.$$

The same estimate holds for $\tilde{\mu}_n$ uniformly in n.

Following the reasoning from [2] it is easy to show that the sequence $\{\mu_n\}$ is tight and satisfies the same a priori estimate. Thus, we can pass to a subsequence $\{\mu_{n'}\}$ which weakly converges to a measure μ. For the sake of simplicity this subsequence will be denoted by $\{\mu_n\}$ again. The limiting measure μ satisfies

$$\sup_{k \in \mathbb{Z}} \int \exp(\lambda |x_k|^N) \ d\mu < \infty, \tag{15}$$

moreover,

$$\sup_n \sup_{k\in\mathbb{Z}} \int \exp(\lambda |x_k|^N)\ d\mu_n < \infty. \tag{16}$$

Let us estimate the relative entropy. We note that μ_n and μ_m are related in the following way:

$$\frac{e^Z \mu_n}{\int e^Z d\mu_n} = \mu_m \times \nu_{m,n},$$

where $Z = -W(x_m, x_{-m}) + W(x_m, x_{m+1}) + W(x_{-m-1}, x_{-m})$, and $\nu_{m,n}$ is a probability measure on $E_{m,n}$. Set: $\mu_{m,n} = \nu_{m,n}$. Then

$$\int \log\Bigl(\frac{d\mu_n}{d(\mu_m \times \mu_{m,n})}\Bigr) d\mu_n = \int (Z - \log \int e^Z\ d\mu_n)\ d\mu_n$$

The desired bound follows immediately from (16) and the assumptions about W.

Take a cyclical shift shift σ_m and a power l. For every μ_n with $n > m+l$ (hence for the limit μ) the measure $\frac{d\mu_n \circ (\sigma_m^l)^{-1}}{d\mu_n}$, $l < m$, has a density $e^{u_{m,l}}$ with

$$\begin{aligned} u_{m,l} &= W(x_{m-l+1}, x_{m-l} + W(x_{-m}, x_{-m+1}) + W(x_{m-l+1}, x_{m-l}) \\ &\quad - W(x_{m+1}, x_{m-l}) - W(x_{-m}, x_m) - W(x_{m-l+1}, x_{-m-1}). \end{aligned}$$

Assumption 3) follows from the same bounds on W and (16).

In order to prove assumption 4) we note that

$$\frac{\Bigl[e^{W(x_n,x_{n+1}) + W(x_{-n-1},x_{-n})} \cdot \mu\Bigr] \circ P_n^{-1}}{\int e^{W(x_n,x_{n+1}) + W(x_{-n-1},x_{-n})} d\mu} = \frac{e^{W(x_n,x_{-n})} \cdot \mu_n}{\int e^{W(x_n,x_{-n})}\ d\mu_n}.$$

The normalizing constants can be easily estimated with the help of a priori bounds for μ and μ_n. Applying assumptions on W one can easily get that

$$\frac{d\mu_n}{d\mu \circ P_n^{-1}} \le C_1 e^{C_2(|x_n|^{N-1} + |x_{-n}|^{N-1})}$$

where C_1, C_2 do not depend on n. Hence, assumption 4) follows immediately from (16). $\square$

References

[1] Ambrosio L., Gigli N., Savaré G., Gradient flows in metric spaces and in the Wasserstein spaces of probability measures, Birkhäuser, 2008.

[2] Albeverio S., Kondratiev Yu.G., Röckner M., Tsikalenko T.V., A priori estimates for symmetrizing measures and their applications to Gibbs states, Journ. of Func. Anal., 171, 2000, 366–400.

[3] Bogachev V.I., Measure theory. V. 1,2. Springer, Berlin – New York, 2007.

[4] Bogachev V.I., Kolesnikov A.V., On the Monge–Ampère equation in infinite dimensions, Infin. Dimen. Anal. Quantum Probab. Related Topics, 8(4), 2005, 547–572.

[5] Bogachev V.I., Kolesnikov A.V., Sobolev regularity for the Monge–Ampere equation in the Wiener space. arXiv: 1110.1822. (to appear in Kyoto Jour. Math.).

[6] Bogachev V.I., Kolesnikov A.V., The Monge–Kantorovich problem: achievements, connections, and perspectives, Russian Mathematical Surveys, 67(5), 2012, 785–890.

[7] Cavalletti F., The Monge problem in Wiener Space. Calcul. of Var.and PDE's , 45(1-2), 2012, 101–124

[8] Decreusefond L., Wasserstein distance on configuration space, Potential Anal., 28(3), 2008, 283–300.

[9] Fang S., Nolot V., Sobolev estimates for optimal transport maps on Gaussian spaces, arxiv: 1207.4907.

[10] Fang S., Shao J. Optimal transport maps for Monge–Kantorovich problem on loop groups. Journ. of Func. Anal., 248(1), 2007, 225–257.
[11] Feyel D., Üstünel A.S., Monge–Kantorovich measure transportation and Monge–Ampère equation on Wiener space. Prob. Theory and Related Fields, 128, 2004, 347–385.
[12] Kolesnikov A.V., Convexity inequalities and optimal transport of infinite-dimensional measures, J. Math. Pures Appl. (9), 83(11), 2004, 1373–1404.
[13] Kolesnikov A.V., On Sobolev regularity of mass transport and transportation inequalities, Theory of Probability and its Applications, to appear. Translated from Teor. Veroyatnost. i Primenen., 57(2), 2012, 296–321.
[14] Kolesnikov A.V., Mass transportation and contractions, MIPT Proc. 2(4), 2010, 90–99.
[15] Kolesnikov A.V., Röckner M., On transport equation in infinite dimensions (work in progress).
[16] Villani C., Topics in optimal transportation, Amer. Math. Soc. Providence, Rhode Island, 2003.

Higher School of Economics, Moscow, Russia
E-mail address: Sascha77@mail.ru

Higher School of Economics, Moscow, Russia
E-mail address: zaev.da@gmail.com